Published in final edited form as:

J Mol Biol. 2007 November 9; 373(5): 1361–1373.

Predicting Allosteric Communication in Myosin via a Pathway of Conserved Residues

Susan Tang^{1,†}, Jung-Chi Liao^{1,†}, Alexander R. Dunn², Russ B. Altman¹, James A. Spudich², and Jeanette P. Schmidt^{1,*}

¹ Department of Bioengineering, Stanford University, Stanford, CA 94305, USA

² Department of Biochemistry, Stanford University Medical Center, Stanford, CA 94305, USA

Summary

We present a computational method that predicts a pathway of residues that mediate protein allosteric communication. The pathway is predicted using only a combination of distance constraints between contiguous residues and evolutionary data. We applied this analysis to find pathways of conserved residues connecting the myosin ATP binding site to the lever arm. These pathway residues may mediate the allosteric communication that couples ATP hydrolysis to the lever arm recovery stroke. Having examined pre-stroke conformations of *Dictyostelium*, scallop, and chicken myosin II as well as *Dictyostelium* myosin I, we observed a conserved pathway traversing switch II and the relay helix, which is consistent with the understood need for allosteric communication in this conformation. We also examined post-rigor and rigor conformations across several myosin species. Although initial residues of these paths are more heterogeneous, all but one of these paths traverse a consistent set of relay helix residues to reach the beginning of the lever arm. We discuss our results in the context of structural elements and reported mutational experiments, which substantiate the significance of the pre-stroke pathways. Our method provides a simple, computationally efficient means of predicting a set of residues that mediate allosteric communication. We provide a refined, downloadable application and source code on <https://simtk.org> to share this tool with the wider community (<https://simtk.org/home/allopathfinder>).

Keywords

allostery; myosin; pathway; conservation; computation

Introduction

Allosteric communication describes one or more events where signal detection at one site in the protein influences other distant sites in the protein by means of conformational changes. A number of important proteins use allostery to mediate function, such as G protein-coupled receptors, hemoglobins, and serine proteases.^{1,2,3} A classic example of allosteric proteins is hemoglobin. Oxygen binding cooperativity is mediated through residues situated in distant secondary structural elements. Hemoglobin is characterized by two equilibrium states, tense (T) and relaxed (R). The two different states are characterized by different affinities for oxygen. According to the alternative sequential model by Koshland, Nemethy and Filmer, ligand

*Corresponding author (jeanette.schmidt@stanford.edu).

[†]S.T. & J.-C.L. made equal contributions to this work.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

binding alters the conformation of the bound subunit, which in turn, impacts ligand affinity in neighboring subunits.⁴ Another well-studied allosteric protein is the bacterial chaperonin GroEL, a multi-subunit protein that harvests energy from ATP to fold proteins. GroEL's internal cavity conformation, responsible for substrate binding, folding, and release, is coupled to ATP binding and hydrolysis at the physically distant apical domain. Significant conformational changes during repetitions of the chaperonin cycle are necessary to correctly fold a protein.⁵

Myosin also uses allostery to mediate its function. Myosin is a motor protein that harvests and transforms ATP energy into movement along actin. X-ray crystallography and biochemical experiments have identified three major myosin motor conformations that correspond to discrete stages of the acto-myosin cycle.⁶ The structure of the *pre-stroke* state is characterized by hydrolyzed ATP in the catalytic site, an open 50 kDa cleft between the catalytic site and the actin binding site, and the lever arm in its cocked, or “up” position. During the power stroke the lever arm swings to its “down” position.

Release of phosphate and ADP results in the *rigor state*, characterized by a closed 50kDa cleft, strong binding to actin, and the lever arm in the “down” position. When the catalytic site is again occupied by ATP, myosin releases from actin and the molecule reaches its *post-rigor* conformation with its lever arm still in the “down” position and an open cleft. We examine the allosteric communication between the ATPase site and the lever arm for these conformations.

Although numerous myosin classes are responsible for diverse cellular functions, the basic power-stroke mechanism is believed to be universal. As expected, the myosin motor domain shows significantly higher conservation than the myosin tail, which encodes cargo specification and/or localization.⁷ Given the conserved catalytic mechanism across different myosin classes and species, we hypothesized that a pathway of allosteric communication from the ATP binding to the beginning of the lever arm could be identified through a set of highly conserved and physically proximate intermediate residues. Thus, we sought to compute a pathway of residues in each conformation of myosin using a combination of distance constraints between adjacent path residues and evolutionary data. While allosteric communication through a set of physically proximate residues is one possible mechanism, other mechanisms have been proposed. As the results presented here will show, it is likely that for the communication from the ATP binding to the lever arm, allostery may indeed occur through a set of physically proximate residues. The premise of our approach is consistent with prior work proposing that structural motifs within allosteric proteins can manifest themselves in an evolutionarily conserved sparse network of residue interactions.⁸

Cope et al⁹ aligned 82 myosin motor domain sequences and identified 131 highly conserved (> 90% conservation) residues dispersed throughout the motor domain. Due to the identification of new myosin sequences since their work, we were able to compile a larger multiple sequence alignment (158 sequences) and use this alignment to evaluate residue conservation. The multiple sequence alignment is available under downloads at <https://simtk.org/home/allopathfinder>. Using physical proximity and conservation we found nearly identical pathways in all pre-stroke structures examined, which included structures for scallop, *Dictyostelium*, and chicken myosin II, as well as *Dictyostelium* myosin I. The beginning portions of the rigor and post-rigor paths are more heterogeneous, but the canonical pre-stroke path is nearly matched thereafter. Furthermore, we discuss our results in light of previously reported experimental mutation effects. The proposed paths residues, which were identified using a combination of distance constraints between adjacent residues and evolutionary data, are consistent with previous studies that identify critical residues by genetic and mutagenesis approaches.

Our method is computationally inexpensive and can be applied to other allosteric proteins after constructing a sufficiently informative multiple sequence alignment. A generalized and easy-to-use application of this method has been made available at <https://Simtk.org/home/allopathfinder>.

Results

Multiple Sequence Alignment and Conservation Weights

We constructed a multiple sequence alignment (MSA) of 158 myosin sequences from 33 different species. The MSA includes five reference sequences corresponding to the motor domains from *Dictyostelium* myosin II, scallop myosin II, chicken smooth muscle myosin II, *Dictyostelium* myosin I, and chicken myosin V. We obtained appropriate PDB structures for these proteins, some bound to various nucleotide analogs, such that multiple examples were included in each of the three major conformations: pre-stroke, rigor and post-rigor. These conformations are classified according to the positions of their lever arm and cleft: pre-stroke (lever arm up, cleft open), rigor (lever arm down, cleft closed), and post-rigor (lever arm down, cleft open). A conservation weight was computed for each column in the MSA that contained a residue from any of the reference structures.

Graph Network

To express potential residue interactions in a given structure, we constructed a graph representation. The graph contains a node for each residue in the structure and an edge between residues whose minimal distance does not exceed 3.3 Å, a distance chosen to reflect an acceptable distance for potential interactions, while minimizing spurious interactions (see Methods). We compared the paths obtained at a cutoff of 3.3 Å with those obtained at 3.4 Å and 3.5 Å and observed that they are very similar. We set the conservation weight of each directed edge to the conservation of the destination residue.

Allosteric Communication from the Catalytic Site

We postulated the source of allosteric communication to be γ -phosphate of the ATP catalytic site. The destination was identified as a region corresponding to the beginning of the lever arm. The precise amino acid residue assigned to be the actual destination was chosen by first clustering the *Dictyostelium* pre-stroke conformation residues based on physical proximity, and subsequently applying a metric to select for paths with high conservation as described below.

Although physically composed of a group of residues, the beginning of the lever arm needs to be concisely represented by a single surface residue for the purpose of our path calculation. To select the appropriate surface residue as our path endpoint, we first clustered the 389 surface residues in the *Dictyostelium* myosin II pre-stroke structure. We then ranked the conservation of paths ending at those residues using our path efficiency ratio (PE), computed as the Cartesian distance between start and end points divided by the sum of path conservation weights. Surface residues were ranked in order of decreasing PE to identify paths with high conservation. We examined whether high-scoring surface residues, individually conferring paths with high conservation per angstrom, converged to a particular region within the structure. Strikingly, although the size of the cluster of interest (containing lever arm residues) varied significantly across various clustering runs, 12 of the 25 residues with the best PE were assigned to the same cluster in over 75% of the clustering runs. These residues, E492, K496, K498, N500, I504, P693, R695, R738, F739, I741, T742, and K743, are situated near or at the beginning of the lever arm. Residues E492, K496 and K498 are part of the relay helix close to the lever arm. Spatial distribution and clustering assignments of the best 25 residues in terms of PE are shown

for one run at $K = 15$ (see Methods) (Figure 1). We selected residue T742 for further analysis as our path endpoint because it had the highest PE of residues mapping to this region.

Shortest Path Calculation for pre-stroke *Dictyostelium* myosin II

All matrices (Identity, Miyata, WAC) yielded two shortest paths with similar trends for the *Dictyostelium* myosin II pre-stroke structure (1VOM),¹⁰ with only minor differences at the residue level (Figure 2(a)). Residue substitutions as well as insertion/deletion of 1 or 2 residues can be detected when comparing the paths. Such differences resulted in the highest-residue path having 12 residues (Miyata) and the least-residue path having 10 residues (WAC & Identity). We chose the Miyata matrix for all further analysis due to its wide acceptance and application.

One of the two identified Miyata paths traverses three conserved motifs shared in all myosins: P-loop (Residues 179–186), switch II (Residues 454–459), and the relay helix (Residues 466–518) (Figure 2(b)).¹¹ The other path skips the P-loop and instead traverses an additional switch II residue but is otherwise identical.

All paths comparisons that follow involve different reference sequences and structures. To facilitate path comparisons, residues computed for each structure were mapped back to residues in *Dictyostelium* myosin II via the MSA. Thus, the residue numbers on the paths presented hereafter have been converted to the residue numbering in *Dictyostelium* myosin II unless otherwise noted.

Shortest Paths in other Pre-stroke Structures

Figure 3 shows the predicted paths of allosteric communications for six pre-stroke structures (teal), seven post-rigor structures (green), and two rigor structures (grey). As is frequently the case, the PDB files corresponding to these structures have missing side chains, which could affect the validity of certain edges and alter the paths. To make sure our results are not affected, we completed all structures with Swiss-Pdb Viewer¹² to fill in all missing side chains. We verified that the paths for these reconstructed structures were in 100% agreement with the paths found using the unmodified PDB files, indicating that the inclusion or absence of the rebuilt residues does not alter our results. The paths of all pre-stroke and post-rigor structures were structurally aligned using MultiProt¹³ and shown in Figure 4.

Comparison of paths in the pre-stroke structures of scallop myosin II (1QVI14) and *Dictyostelium* myosin II show that they are nearly identical, and that the differences can be attributed to a minor structural deviation. Investigation of additional pre-stroke structures reiterate this trend, including *Dictyostelium* myosin I (1LKX)¹⁵, chicken smooth muscle myosin II (1BR2)¹⁶, as well as two *Dictyostelium* myosin II pre-stroke structures 1W9J (S456Y, bound with MgADP-AlF₄) and 1YV3¹⁷ (bound to blebbistatin and ADP VO₄) (Figure 3, 4(a), 4(b)). Paths of all six pre-stroke structures pass through residue F482 in the relay helix. After residue F482, all paths follow the relay helix and closely mirror one another with minor divergences.

As observed above for the *Dictyostelium* myosin II 1VOM structure, paths before residue F482 seem to have two options for each of the structures. One option is to jump directly to switch II (almost always to residue G457 and then followed almost always by F458, with the exception of 1BR2, whose paths go directly to F458). The other option is to traverse a single residue in the P-loop followed by a single residue in switch II (typically the pair (S181, F458)). Exceptions are the *Dictyostelium* myosin I (1LKX) structure, which has no optional jump from the binding site directly to switch II and for which all paths start with P-loop residue S181, and 1QVI which prefers the direct jump to the relay helix (G457).

Five of the six structures go through residue N475 in the relay helix, either coming directly from residue F458 in switch II or by going through residue I471 first. The *Dictyostelium* myosin I (1LKX structure) is again slightly different in one of its paths, where due to a large spatial gap between residue F458 and N475, the path traverses three additional residues in the relay helix before arriving at N475. In summary, all pre-stroke paths traverse 0–1 residue in the P-loop (S181), 1–2 residues (among G457, F458) in switch II, and 8–10 residues in the relay helix.

Paths Calculated for Post-Rigor Configurations

We applied a similar shortest path calculation to the corresponding scallop myosin II apo post-rigor conformation: 1KK7,¹⁸ chicken myosin V post-rigor with ADP-BeF3 bound (1W7J),¹⁹ as well as *Dictyostelium* myosin II post-rigor structures: 1MMD,²⁰ which is bound to ADP-BeF3; 1FMW,²¹ ATP; 1FMV,²¹ apo; 1W9I, S456Y mutation, ADP-BeF3; and 1MMA,²² ADP (Figure 3, 4(c), 4(d)). Three of these structures have paths adhering closely to pre-stroke paths (1MMD, 1KK7 and 1FMV) traversing K185 and sometimes 1–2 additional residues (among G179, E180, S181) in the P-loop, followed by 1–4 residues in switch II (among D454 through E459), and then 7–10 residues in the relay helix. In contrast, the ADP-bound 1MMA structure exhibits a very different path. The path in this post-rigor structure starts in the P-loop, weaves through numerous residues in switch I, goes through a few of the common residues in switch II, and then traverses several residues in the 670–700 (SH1–SH2) region before arriving at residue T742 (Figure 5). We note that this divergence is not an artifact of the path with lowest weight, but is due to structural differences, (Figure 5).

We note that the pre-stroke structures tend to have a few alternate near-optimal paths, especially before reaching the relay helix, which is only the case for 1KK7 among the post-stroke structures (see Discussion).

Paths in Additional Rigor and Post-Rigor Structures

We further applied our method to the additional rigor structures *Dictyostelium* apo myosin II rigor (2AKA)²³, and chicken apo myosin V rigor (1OE9).²⁴ When paths from all structures (including the previously discussed *Dictyostelium* myosin II and scallop myosin II structures) are aligned (Figure 3), we found that seven of thirteen matching residues in the pre-stroke paths are also present in most post-rigor and rigor paths. However the post-rigor and rigor paths appear to show somewhat more discrepancy at the beginning of their paths (Figure 4(c), (d)). Some paths traverse several P-loop and switch I residues before visiting the block of frequently observed residues. The Myosin V rigor structure (1OE9) has 10 paths, and the *Dictyostelium* apo myosin II rigor structure (2AKA) has 12 paths with nearly equivalent conservation scores, that differ significantly before residue N475 from the paths for other structures but converge thereon after. For 2AKA a few alternate paths continue to diverge from the general template up to residue F482. When increasing the cutoff distance to 3.4 Å, the paths remain similar but some shortcuts are observed. In particular, at distance 3.4 Å the paths for 1QVI show the shortcuts that the path for 1YV3 exhibit at cutoff 3.3 Å, preferring the route through G457 over the paths through S181. For almost all structures, paths take one shortcut at higher cutoff values towards the end of the paths, skipping 1 residue per path for 3.4 Å and 2 residues per path for a cutoff distance of 3.5 Å.

Discussion

We have demonstrated a very simple and computationally inexpensive method for determining a putative set of residues involved in protein allostery. Our method requires (1) sequences (sufficient in number to obtain a meaningful multiple sequence alignment) from different species for proteins where similar allostery is suspected, (2) the conformation in which the

allosteric pathway is assumed to exist, and (3) the starting and ending positions of the allosteric path. It is generally helpful to repeat the computation on several comparable structures represented by sequences in the MSA to establish the significance of paths obtained.

Using residue conservation as inferred from the multiple sequence alignment as well as distance constraints described by the conformational structures, we predicted a set of intermediate residues that physically couple subtle changes from the nucleotide binding site to large lever arm motions. The identified paths are conserved among pre-stroke structures from different species and across two myosin classes. Furthermore, analysis of all post-rigor and rigor conformations produced paths (with the exception of the 1MMA post-rigor structure) aligning remarkably well in the relay helix, but showing significant variation at the beginning of the paths.

The similarity of paths in all pre-stroke structures examined (Figure 3, 4(a), 4(b)) reflects the robustness of our method. In particular, remarkable conservation was observed between paths in the pre-stroke structure 1VOM (*Dictyostelium* myosin II) and paths in the pre-stroke structure 1QVI of scallop myosin, with very minor differences attributed to subtle structural differences. These sequences exhibit ~45% identity within the motor domain. In contrast, path correspondence is sometimes markedly lower for different conformations of the *same* protein: the 1VOM path has no similarity to the 1MMA path, and while the paths from 1QVI and 1KK7 share similarity, the 1QVI paths are closer to the 1VOM paths than to the 1KK7 paths. Paths differences in the post-rigor (Figure 4(c), 4(d)), and rigor structures may demonstrate the existence of alternate allosteric communication paths during other states in the catalytic cycle. The unusual paths in 1MMA may lack biological significance, since this ADP-bound structures represents a state that is not present in the physiologically relevant catalytic cycle. For completeness we also analyzed the paths for 1B7T another ADP-bound structures not present in the physiologically relevant catalytic cycle and observed a path that is both different from the 1MMA path as well as the “usual” path.

A number of computational methods, including normal mode analysis,^{11,25,26} energy minimization algorithms,²⁷ and molecular dynamics simulations,²⁸ have been used to identify putative coupling mechanisms between the ATP binding site and the lever arm. Stastical coupling analysis (SCA),^{8,29} a method based on sequence information, has also been used to identify coupling residues of several other proteins. Recently, Yu *et al.* implemented an approach that combines molecular dynamics, normal mode analysis, hinge analysis, and SCA to propose critical residues for coupling.^{30,31} Based on normal mode analysis, Zheng *et al.* identified hinge residues which correspond to high correlation.²⁵ They found important residues in the ATP binding site, the actin binding site, the converter, the relay helix, and the SH1 helix. They further advanced the method by incorporating evolution information as perturbation strengths and obtained a list of hotspot residues for myosin.²⁶ The SCA method developed by Ranganathan and coworkers provides coevolution analysis and was used to identify allosteric coupling residues.^{8,29} Mechanochemical studies by Cui and co-workers integrated the SCA method (for identifying coupling residues) with normal mode analysis to find hinge residues, and further incorporated the interaction analysis using molecular dynamics simulation to explore the key residues of the allosteric network for myosin.³¹ Our approach, which aimed to identify physically interacting residues forming paths from the ATP binding site to the lever arm (or in general from point A to point B), provides a different venue for interpreting allosteric interactions. The method is remarkably simple and shows that in some cases the conserved mechanism driven by allostery can be explained by a sequence of conserved residues, and gives results consistent with experimental studies as well as other computational studies such as those discussed above.

Signal Propagation from Catalytic Site to Lever Arm

According to the swinging crossbridge mechanism,³² allosteric communication initiates from the ATP binding site of the catalytic domain and travels to the lever arm on the catalytic domain surface during the lever arm cocking step in the myosin cycle. However, in order to avoid the bias inherent in searching for paths between only these two sites, we ranked all putative paths of allosteric communication between the ATP catalytic site and residues at the catalytic domain surface for 1VOM. We proceeded to examine whether high scoring paths disproportionately terminate at the lever arm. Reasoning that potentially interesting paths would travel through conserved residues, we defined a path efficiency metric (PE), which reflects “conservation per angstrom traveled”. Path efficiency (PE) scores were computed for the best (lowest weight) path generated between the ATP catalytic site and each surface residue. We grouped surface residues using K-means clustering and discovered that surface residues with the best PE values demonstrate scattered distribution except for a collection of residues corresponding to the beginning of the lever arm. This collection of residues group to the same cluster in a majority of clustering runs. This result substantiates our hypothesis that signals appear to propagate from the catalytic site preferentially to the lever arm. Using PE as a measure, we were then able to rationally select a residue from the beginning of the lever arm for subsequent calculations. While a disproportionate number of paths with highest PE score ended at the “lever arm cluster”, we note that the path with the highest PE score in 1VOM ends in A286, a residue distant from the beginning of the lever arm and not known to have any functional importance. When we applied a similar analysis to the scallop pre-stroke structure, we found that the residue corresponding to A286 is buried and therefore cannot serve as a path’s final residue. This indicates, not surprisingly, that not all end points of paths with high PE scores have functional significance, and that a path needs to be conserved across species or structures to be valid. Together, these results suggest that the combination of PE ratio calculation and cluster analysis may be useful in predicting end points for long-range allosteric pathways.

The pre-stroke pathway we observe is consistent with previous descriptions of the recovery stroke.^{10,16,27,32} Myosin can be modeled as four subdomains that are linked by three joints. According to this model, rigid-body motion of the subdomains and concerted changes in the single-stranded joints are required to proceed from one conformation to another in the cycle.³³ Joints are crucial components because they facilitate information flow between distant, rigid regions of the protein. A major structural difference observed across different conformational states is the relative positioning of the converter domain with respect to the relay helix, which can be modeled as a joint.⁶ A piston-like motion of the relay helix is thought to be responsible for directing converter and hence lever arm motion.³⁴ The repositioning of the relay helix is in turn thought to result from the movement of switch II.¹⁹ Switch II plays a major role in sensing the state of the nucleotide γ -phosphate.^{16,35,36} Crystal structures using ATP analogs indicate that the presence of ADP and phosphate in the catalytic site causes switch II to gravitate towards the hydrolyzed phosphate, which in turn triggers movement of the relay helix and converter and results in the lever arm transition to the cocked, “upstroke” position. The common pathways we observe in the pre-stroke structures based solely on evolutionary data and distance constraints agree with the above mechanistic studies. This agreement demonstrates the robustness of our method, and motivates the detailed examination of the path residues found for pre-stroke, post-rigor, and rigor conformations, as discussed below.

Path conservation and divergence

The fundamental similarity of the pre-stroke pathways across three species and two myosin classes is striking and its highly conserved nature may reflect a single, well-defined state within the catalytic cycle. A similar overall path is also observed in the majority of post-rigor structures (Figure 3). In particular, despite differences in the P-loop and/or switch I, these paths use a nearly identical sequence of residues in the relay helix to reach the beginning of the lever arm.

This similarity suggests that a common mechanism of allosteric communication may operate in both the post-rigor and pre-stroke states, and is consistent with the interpretation that the transition state analogs trap myosin in a state that is poised at the beginning of the recovery stroke. The greater diversity in post-rigor paths likely stem from the variety of nucleotide states and switch I and II positions that are compatible with the overall post-rigor conformation.

The nucleotide-free structures 1FMV and 1KK7 generate paths that are similar to the pre-stroke paths and that are also similar to the ATP-analog bound post-rigor structures (e.g. 1MMD). Although we classify the scallop nucleotide-free structure 1KK7 as “post-rigor” based on its mostly open 50 kDa cleft, it has been described as “near-rigor” in the literature.¹⁸ Similarly, nucleotide-free *Dictyostelium* myosin II has been crystallized in both a rigor-like (2AKA) and post-rigor conformations (1FMV). These observations suggest that in the absence of actin and nucleotide, the rigor and post-rigor conformations are energetically comparable for *Dictyostelium* myosin II.

The P-loop, switch I, and switch II

In the pre-stroke paths, the γ -phosphate analog is coupled to switch II directly through a hydrogen bond to residue G457 (F458 in chicken myosin II), or indirectly through the P-loop residue S181. The direct phosphate-switch II interaction has been noted previously, and has been hypothesized to play a key role in establishing the structural transformations that lead to the recovery stroke.¹⁰

The initial portions of the post-rigor paths show greater diversity (Figure 3, 4(c), 4(d)). All paths begin in the P-loop at residue K185, before connecting to switch II via the P-loop residues G179, E180, S181 or G182 (we note that paths in apo structures were *defined* to begin at residue K185). In the pre-stroke conformation, S181 offers a connection from the γ -phosphate to switch II through a main-chain hydrogen bond, but only because switch II has moved close to the ATP binding site. A gap between the P-loop and switch II opens in the post-rigor structures, necessitating a more circuitous route through K185 and G179/E180 before continuing on to switch II and the relay helix.

The P-loop forms a key part of the ATP binding site, and is very highly conserved among myosins and is structurally conserved among diverse NTPases, including G proteins, kinesin, and ATP synthase.^{37,38,39} Despite its obviously central role in catalysis, the P-loop has been subject to relatively little scrutiny in myosin, probably because its conformation is static in nucleotide-bound structures (though not in the apo state, see below). Our pathway calculations highlight how the P-loop mediates *indirect* contacts between the nucleotide and switch II, and may thus play a more active role in myosin allostery than is generally appreciated.

Paths computed for the nucleotide-free, rigor-like structures (2AKA and 1OE9) include long digressions through either the P-loop or switch I, before converging to the common pathway through the relay helix. Examination of both of these structures shows a large ($> 5 \text{ \AA}$) gap between residues 181–185 and the closest residue in switch II, explaining the insertion of switch I residues (1OE9) or the addition of several P-loop residues (2AKA). This structural rearrangement is related to the rearrangements in the nucleotide binding pocket.²⁴ In the rigor-like myosin V structure, residue S181 hydrogen bonds to residue N233 in switch I. Nucleotide binding necessitates a structural rearrangement wherein this interaction is exchanged for hydrogen bonds to the γ -phosphate. One possible interpretation of our result is that ATP binding restores direct communication between the P loop and switch II, which could be functionally related to the opening of the 50 kDa cleft and the eventual detachment from actin. Although the beginning residues of the rigor-like paths are divergent, strong path conservation is again observed in the relay helix. This observation supports the centrality of the relay helix in mediating communication between the lever arm and ATPase site.

Known Pre-stroke Pathway Mutants

Four pathway residues (F458, F482, F487, I499) are known to result in functional defects when mutated. The mutant F458A, present in all pre stroke structure with the exception of 1QVI, shows a nearly complete loss of actin-activated ATPase activity.⁴⁰ F482 is present in every one of our pre-stroke paths. The mutation F482A dramatically slows ATP hydrolysis even though the residue is distant from the active site.⁴¹ F487 is present in all but one of the pre-stroke paths and the mutation F487A increases actin affinity in the presence of ATP, thus disrupting the normal release from the actin filament in the post-rigor state.⁴² While I499 is present on very few of our paths, all of our paths (with the exception of 1MMA) go through either E497 or G498. The distance between these paths residues and I499 in our twelve structures that do not go through I499 ranges from 1.3Å to 3.4Å (with an average of 2.7Å). The F499A mutant hydrolyzes ATP with close to normal kinetic parameters, but does not produce a functional lever arm swing.⁴³ These experimental results underscore the importance of at least some of the pre-stroke path residues in mediating allosteric communication. Based on the agreement between our work and previous experimental results, we speculate that other highly represented path residues, for instance S181, K185, N475 and Q479, are excellent candidates for further mechanistic studies.

Other relevant known mutations

Several known mutations to residues in the SH1 helix alter myosin catalytic activity, and in particular the coupling between the lever arm, ATP binding site, and actin binding interface. These residues often contact pathway residues. For instance, the *Dictyostelium* myosin II mutation F692A uncouples ATPase activity from the lever arm swing.⁴³ F692 hydrogen-bonds to the side chain of pathway residue Y494, and is in close proximity to the pathway residue I499.

The mutations G680V, G680A and G680F result in an abolishment of function due to slowed phosphate release and a structural alterations in the actomyosin interface.^{44,45} G680 contacts F482, a highly conserved pathway residue, which is itself sensitive to mutation (see above). Remarkably, the combined mutations G680F, F482A almost completely restores function, suggesting that contacts between the relay and SH1 helices are essential for mediating long-range allostery.⁴¹ Similarly, a disulfide bond introduced using the double mutant I499C and R738C has no effect on catalytic parameters, again indicating that this interface is conserved during the pre- to post-stroke structural transition.⁴⁶

Taken together, mutational studies suggest that contacts between pathway residues and the SH1 helix are functionally critical, thus further underscoring the functional validity of the pathway identification. The existence of these SH1 helix mutations, which lie near but not on our identified pathway, also demonstrates that the algorithm used here does not capture the full complexity of myosin structural dynamics, but rather focuses on identifying residues critical for the allosteric communication.

Conclusion

Allosteric communication pathways are often invoked when discussing the biological significance of protein structures. These pathways are usually determined using personal structural intuition, with reference to the available biochemical and functional data. Our method provides a simple, computationally efficient means of predicting a set of residues using evolutionary constraints and only one three-dimensional structure. We provide a refined, downloadable application at <http://simtk.org/home/allopathfinder> and source code to share the tool with a wider community.

Methods

Multiple Sequence Alignment

Two sources were used for the accumulation of a comprehensive set of protein sequences for myosin II motor domain. We used a large set of myosin sequences previously reported in a multiple sequence alignment (MSA) format [<http://www.mrc-lmb.cam.ac.uk/myosin/trees/colour.html>, May 2000]. This set of sequences was enriched with additional published and aligned myosin sequences.⁴⁷ We extracted corresponding myosin sequences (including any updates to sequences used in the original MSA) from other databases including NCBI, GenDB, TIGR, and DOE-JGR. The multiple sequence alignment of the 158 identified sequences was constructed using ClustalW, using the PAM250 substitution matrix. We made sure to include all reference sequences that correspond to myosin conformations that participate in our subsequent analysis in the MSA.

Conservation Weights

Our goal was to assign a conservation weight to each column of the multiple sequence alignment with the following properties: all weights are positive and small weights reflect high conservation. We used the sum of pairs method, where substitution scores for each (i, j) residue pair in the column are aggregated to assign a raw conservation weight using three different conservation matrices as described below. To avoid zero weights (consistent with the assumption that some cost is associated with traveling through any residue) we initialized all weights to 0.001 and computed weights using 10^{-2} precision to produce optimal paths. For near-optimal paths we relaxed the precision to 10^{-1} while using an initialization of 0.01.

We chose three matrices for the analysis. The simplest matrix is the identity matrix with a substitution score of one for a pair of matching amino acids and zero otherwise. The Miyata matrix reflects the idea that the frequency of amino acid substitutions decreases linearly with increase of physical-chemical differences between the amino acid pair.⁴⁸ The pairwise scores (distances) between amino acids using the Miyata matrix reflect differences in both physical attribute (volume) and chemical attribute (polarity) between the pair. The Wei-Altman-Chang (WAC) matrix examines physical-chemical and structural differences between environments of amino acid pairs to assign a score to the substitution.⁴⁹

Graph Network

To find the most probable set of residues for allosteric communication, the residues in a protein structure were represented by a directed, weighted graph. Each vertex in the graph represents a residue, and each weighted edge represents potential interaction between a residue pair. We draw an edge (u, v) between residues u and v if the distance between these residues in the structure falls below a threshold value chosen to reflect the maximum acceptable distance for a potential interaction. In our graph, a bi-directional edge was added between any two residues whose minimal contact distance (distance between any atom in residue u and any atom in residue v) was below 3.3 Å. Since we expect myosin allosteric communication to propagate through a network of hydrogen bonds and Van der Waals interactions, a distance constraint slightly more generous than the hydrogen bond length (3 Å) was selected. We also increased the distance to 3.4 Å and 3.5 Å and observed no significant differences. Directed edge weights are derived from conservation of the destination residue. The weight of edge (u, v) is defined to be the conservation weight of residue v, given that the likelihood of traveling from u to v should be highly correlated to the conservation of v under our model. The weight associated with traversing an edge from u to v is hence different from the weight associated with traversing the same edge from v to u. Using such a framework, the problem of finding the most conserved set of residues is simplified to finding the least weight (or shortest) path in the graph network between a source and sink node.

Dijkstra's Algorithm

The weight associated with a path from v_i to v_f is computed by summing the weights of edges connecting the path from v_i to v_f . To find the least weight path from v_i to v_f , we used Dijkstra's algorithm. Dijkstra's algorithm is an efficient algorithm that has been shown to find the path of least weight (defined as the sum of weights on the path) from a specified initial vertex to all other vertices in a graph provided that each edge is associated with a positive weight.⁵⁰ When given an initial vertex v_i , Dijkstra's algorithm will produce a shortest path from v_i to all other reachable vertices in the graph. The set of shortest paths can be extracted from the graph in order of increasing path weight. Since the algorithm requires all weights to be positive, paths whose traversed nodes are characterized by high conservation will be preferentially extracted.

Allosteric Communication from the Catalytic Site

To apply Dijkstra's algorithm to our problem, we need to define an initial and final position for the communication pathway within a protein structure. The initial position marks the source of allosteric communication, and is unequivocally represented by γ -phosphate in the catalytic site. It is understood that ATP hydrolysis energy is stored in β - γ -phosphoanhydride bond and therefore structural change must propagate from there. Depending on the nucleotide presence at the catalytic site, the initiation of the path was selected as either γ -phosphate analog (ADP- γ -phosphate analog is present), β -phosphate of ADP (only ADP is present), or the equivalent of S181 in *Dictyostelium* myosin II (apo state; no nucleotide). Although the basic mechanism and ATPase cycle of myosin strongly suggests that structural changes must propagate from the ATP binding site to the lever arm, we also searched for the most probable direction of allosteric communication without using our prior knowledge of function and mechanism.

Residues in 1VOM (*Dictyostelium* pre-stroke) were first classified as either "buried" or "surface" using NACCESS.⁵¹ The relative surface accessibility (RSA) for each amino acid is calculated by dividing the solvent accessible surface area by the maximal accessible surface area of the amino acid. A residue was labeled as surface and hence a candidate for a path end residue, if its RSA > 25 % and buried otherwise.

We grouped residues using K-means clustering of spatial coordinates to make regional inferences from individual surface residues. K-means clustering was performed on all 730 defined PDB residues of 1VOM, using 3D coordinates of the C α atom to represent each residue and Euclidian distance from a cluster center to group the residues. Residues in spatial proximity are grouped together in such a way that K or fewer nonempty residue clusters are produced. Our implementation of K-means clustering terminates when data assignments to clusters converge, and each residue is closest to the mean point (centroid) of its current cluster than to the centroid of any other cluster. Three different values of K were explored (K = 10, 15, and 20). Since random initial cluster assignments can impact the final solution, we ran the algorithm three times for each value of K to acknowledge variability in results.

We defined an efficiency metric as $PE = PD/PW$, where PD represents the Cartesian distance between the initial and final positions and PW represents the net path weight. In the context of a path, PE measures the conservation per angstrom traveled, (higher conservation resulting in a higher PE value) thus enabling paths of different lengths to compete fairly in terms of path conservation. Two factors contribute to the path weight: number of residues used and conservation of those residues. We ordered the surface residues in decreasing efficiency, limiting our analysis to residues whose paths encompass a distance of > 20 Å. This distance minima criterion eliminates paths that never leave the ATP binding site. Using earlier clustering assignments, we analyzed endpoints of the best scoring paths and looked for well-represented clusters among them.

T742 of *Dictyostelium* myosin II was identified to have the highest PE among residues in the largest cluster, the cluster containing lever arm residues. The final position for all paths the path was selected to be the residue that corresponds to T742 in *Dictyostelium* myosin II, in our multiple sequence alignment. We reconstructed missing side chains for all structures with Swiss-Pdb Viewer¹².

Shortest Path Calculations

Each path calculation was contingent upon the definition of an initial and final position for the structure being analyzed. For each conformation that we examined (pre-stroke, post-rigor, and rigor), we were able to locate solved structures from two myosin classes and at least two different species. For the pre-stroke conformation, structures were selected from Myosin I and II. For the post-rigor and rigor conformation, structures were selected from Myosin II and V. For all three conformations, we examined structures from at least two of the following: *Dictyostelium*, scallop, and chicken. In particular, we focused on pre-stroke and post-rigor structures from *Dictyostelium* and scallop myosin II for cross-species comparison.

Acknowledgements

This work was supported by the National Institutes of Health through the NIH Roadmap for Medical Research Grant U54 GM072970 (to S.T., J.-C.L., R.B.A, J.A.S. and J.P.S.), the National Institutes of Health Grant GM33289 (J.A.S. and A.R.D.), and the Jane Coffin Childs Memorial Fund (A.R.D.).

References

1. Dang QD, Di Cera E. Residue 225 determines the Na(+)-induced allosteric regulation of catalytic activity in serine proteases. *Proc Natl Acad Sci U S A* 1996;93:10653–6. [PubMed: 8855234]
2. Gether U. Uncovering molecular mechanisms involved in activation of G protein-coupled receptors. *Endocr Rev* 2000;21:90–113. [PubMed: 10696571]
3. Perutz MF. Stereochemistry of cooperative effects in haemoglobin. *Nature* 1970;228:726–39. [PubMed: 5528785]
4. Koshland DE Jr, Nemethy G, Filmer D. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* 1966;5:365–85. [PubMed: 5938952]
5. Keskin O, Bahar I, Flatow D, Covell DG, Jernigan RL. Molecular mechanisms of chaperonin GroEL-GroES function. *Biochemistry* 2002;41:491–501. [PubMed: 11781087]
6. Houdusse A, Szent-Gyorgyi AG, Cohen C. Three conformational states of scallop myosin S1. *Proc Natl Acad Sci U S A* 2000;97:11238–43. [PubMed: 11016966]
7. Cheney RE, Mooseker MS. Unconventional myosins. *Curr Opin Cell Biol* 1992;4:27–35. [PubMed: 1558751]
8. Suel GM, Lockless SW, Wall MA, Ranganathan R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat Struct Biol* 2003;10:59–69. [PubMed: 12483203]
9. Cope MJ, Whisstock J, Rayment I, Kendrick-Jones J. Conservation within the myosin motor domain: implications for structure and function. *Structure* 1996;4:969–87. [PubMed: 8805581]
10. Smith CA, Rayment I. X-ray structure of the magnesium(II).ADP.vanadate complex of the *Dictyostelium discoideum* myosin motor domain to 1.9 Å resolution. *Biochemistry* 1996;35:5404–17. [PubMed: 8611530]
11. Zheng W, Brooks B. Identification of dynamical correlations within the myosin motor domain by the normal mode analysis of an elastic network model. *J Mol Biol* 2005;346:745–59. [PubMed: 15713460]
12. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 1997;18:2714–23. [PubMed: 9504803]
13. Shatsky M, Nussinov R, Wolfson HJ. A method for simultaneous alignment of multiple protein structures. *Proteins* 2004;56:143–56. [PubMed: 15162494]

14. Gourinath S, Himmel DM, Brown JH, Reshetnikova L, Szent-Gyorgyi AG, Cohen C. Crystal structure of scallop Myosin s1 in the pre-power stroke state to 2.6 a resolution: flexibility and function in the head. *Structure* 2003;11:1621–7. [PubMed: 14656445]
15. Kollmar M, Durrwang U, Kliche W, Manstein DJ, Kull FJ. Crystal structure of the motor domain of a class-I myosin. *Embo J* 2002;21:2517–25. [PubMed: 12032065]
16. Dominguez R, Freyzon Y, Trybus KM, Cohen C. Crystal structure of a vertebrate smooth muscle myosin motor domain and its complex with the essential light chain: visualization of the pre-power stroke state. *Cell* 1998;94:559–71. [PubMed: 9741621]
17. Allingham JS, Smith R, Rayment I. The structural basis of blebbistatin inhibition and specificity for myosin II. *Nat Struct Mol Biol* 2005;12:378–9. [PubMed: 15750603]
18. Himmel DM, Gourinath S, Reshetnikova L, Shen Y, Szent-Gyorgyi AG, Cohen C. Crystallographic findings on the internally uncoupled and near-rigor states of myosin: further insights into the mechanics of the motor. *Proc Natl Acad Sci U S A* 2002;99:12645–50. [PubMed: 12297624]
19. Coureux PD, Sweeney HL, Houdusse A. Three myosin V structures delineate essential features of chemo-mechanical transduction. *Embo J* 2004;23:4527–37. [PubMed: 15510214]
20. Fisher AJ, Smith CA, Thoden JB, Smith R, Sutoh K, Holden HM, Rayment I. X-ray structures of the myosin motor domain of Dictyostelium discoideum complexed with MgADP.BeFx and MgADP.AIF₄. *Biochemistry* 1995;34:8960–72. [PubMed: 7619795]
21. Bauer CB, Holden HM, Thoden JB, Smith R, Rayment I. X-ray structures of the apo and MgATP-bound states of Dictyostelium discoideum myosin motor domain. *J Biol Chem* 2000;275:38494–9. [PubMed: 10954715]
22. Gulick AM, Bauer CB, Thoden JB, Rayment I. X-ray structures of the MgADP, MgATPgammaS, and MgAMPPNP complexes of the Dictyostelium discoideum myosin motor domain. *Biochemistry* 1997;36:11619–28. [PubMed: 9305951]
23. Reubold TF, Eschenburg S, Becker A, Leonard M, Schmid SL, Vallee RB, Kull FJ, Manstein DJ. Crystal structure of the GTPase domain of rat dynamin 1. *Proc Natl Acad Sci U S A* 2005;102:13093–8. [PubMed: 16141317]
24. Coureux PD, Wells AL, Menetrey J, Yengo CM, Morris CA, Sweeney HL, Houdusse A. A structural state of the myosin V motor without bound nucleotide. *Nature* 2003;425:419–23. [PubMed: 14508494]
25. Zheng W, Brooks BR, Thirumalai D. Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations. *Proc Natl Acad Sci U S A* 2006;103:7664–9. [PubMed: 16682636]
26. Zheng W, Doniach S. A comparative study of motor-protein motions by using a simple elastic-network model. *Proc Natl Acad Sci U S A* 2003;100:13253–8. [PubMed: 14585932]
27. Fischer S, Windshugel B, Horak D, Holmes KC, Smith JC. Structural mechanism of the recovery stroke in the myosin molecular motor. *Proc Natl Acad Sci U S A* 2005;102:6873–8. [PubMed: 15863618]
28. Koppole S, Smith JC, Fischer S. Simulations of the myosin II motor reveal a nucleotide-state sensing element that controls the recovery stroke. *J Mol Biol* 2006;361:604–16. [PubMed: 16859703]
29. Lockless SW, Ranganathan R. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* 1999;286:295–9. [PubMed: 10514373]
30. Yu H, Ma L, Yang Y, Cui Q. Mechanochemical coupling in the myosin motor domain. I. Insights from equilibrium active-site simulations. *PLoS Computational Biology* 2007;3:e21. [PubMed: 17291159]
31. Yu H, Ma L, Yang Y, Cui Q. Mechanochemical coupling in the myosin motor domain. II. Analysis of critical residues. *PLoS Computational Biology* 2007;3:e23. [PubMed: 17305418]
32. Spudich JA. The myosin swinging cross-bridge model. *Nat Rev Mol Cell Biol* 2001;2:387–92. [PubMed: 11331913]
33. Houdusse A, Kalabokis VN, Himmel D, Szent-Gyorgyi AG, Cohen C. Atomic structure of scallop myosin subfragment S1 complexed with MgADP: a novel conformation of the myosin head. *Cell* 1999;97:459–70. [PubMed: 10338210]
34. Reubold TF, Eschenburg S, Becker A, Kull FJ, Manstein DJ. A structural model for actin-induced nucleotide release in myosin. *Nat Struct Biol* 2003;10:826–30. [PubMed: 14502270]

35. Lawson JD, Pate E, Rayment I, Yount RG. Molecular dynamics analysis of structural factors influencing back door pi release in myosin. *Biophys J* 2004;86:3794–803. [PubMed: 15189875]
36. Murphy CT, Rock RS, Spudich JA. A myosin II mutation uncouples ATPase activity from motility and shortens step size. *Nat Cell Biol* 2001;3:311–5. [PubMed: 11231583]
37. Smith CA, Rayment I. Active site comparisons highlight structural similarities between myosin and other P-loop proteins. *Biophys J* 1996;70:1590–602. [PubMed: 8785318]
38. Kull FJ, Vale RD, Fletterick RJ. The case for a common ancestor: kinesin and myosin motor proteins and G proteins. *J Muscle Res Cell Motil* 1998;19:877–86. [PubMed: 10047987]
39. Goody RS, Hofmann-Goody W. Exchange factors, effectors, GAPs and motor proteins: common thermodynamic and kinetic principles for different functions. *Eur Biophys J* 2002;31:268–74. [PubMed: 12122473]
40. Sasaki N, Shimada T, Sutoh K. Mutational analysis of the switch II loop of Dictyostelium myosin II. *J Biol Chem* 1998;273:20334–40. [PubMed: 9685384]
41. Ito K, Uyeda TQ, Suzuki Y, Sutoh K, Yamamoto K. Requirement of domain-domain interaction for conformational change and functional ATP hydrolysis in myosin. *J Biol Chem* 2003;278:31049–57. [PubMed: 12756255]
42. Tsiavaliaris G, Fujita-Becker S, Batra R, Levitsky DI, Kull FJ, Geeves MA, Manstein DJ. Mutations in the relay loop region result in dominant-negative inhibition of myosin II function in Dictyostelium. *EMBO Rep* 2002;3:1099–105. [PubMed: 12393751]
43. Sasaki N, Ohkura R, Sutoh K. Dictyostelium myosin II mutations that uncouple the converter swing and ATP hydrolysis cycle. *Biochemistry* 2003;42:90–5. [PubMed: 12515542]
44. Batra R, Geeves MA, Manstein DJ. Kinetic analysis of Dictyostelium discoideum myosin motor domains with glycine-to-alanine mutations in the reactive thiol region. *Biochemistry* 1999;38:6126–34. [PubMed: 10320339]
45. Uyeda TQ, Tokuraku K, Kaseda K, Webb MR, Patterson B. Evidence for a novel, strongly bound acto-S1 complex carrying ADP and phosphate stabilized in the G680V mutant of Dictyostelium myosin II. *Biochemistry* 2002;41:9525–34. [PubMed: 12135375]
46. Shih WM, Spudich JA. The myosin relay helix to converter interface remains intact throughout the actomyosin ATPase cycle. *J Biol Chem* 2001;276:19491–4. [PubMed: 11278776]
47. Richards TA, Cavalier-Smith T. Myosin domain evolution and the primary divergence of eukaryotes. *Nature* 2005;436:1113–8. [PubMed: 16121172]
48. Miyata T, Miyazawa S, Yasunaga T. Two types of amino acid substitutions in protein evolution. *J Mol Evol* 1979;12:219–36. [PubMed: 439147]
49. Wei L, Altman RB, Chang JT. Using the radial distributions of physical features to compare amino acid environments and align amino acid sequences. *Pac Symp Biocomput* 1997:465–76. [PubMed: 9390315]
50. Dijkstra EW. A note on two problems in connexion with graphs. *Numerische Mathematik* 1959;V1:269–271.
51. Hubbard, SJ.; Thornton, JM. NACCESS. 'NACCESS'. Department of Biochemistry and Molecular Biology; University College, London: 1993.

Figure 1.

Combined pathway and clustering analysis identifies surface residues with potentially strong allosteric coupling to the ATPase site. Shown are the top 25 high scoring end-point residues (space-filling steric representation) for the *Dictyostelium* pre-stroke structure 1VOM. Twelve non-empty clusters are differentiated by contrasting colors ($K = 15$, see Methods). A green star indicates the origin of signal propagation in the ATPase site. 12 residues (E492, K496, K498, N500, I504, P693, R695, R738, F739, I741, T742, K743) map to the lever arm cluster (magenta). The remaining residues are distributed amongst other clusters. The cluster with the second largest number of high PE residues contains only 4 residues (yellow).

(181 or 457)	458	475	479	483	--	--	487	491	493	--	--	497	742
9	9	9	9	4			5	6	5			8	5
(181 or 457)	458	475	479	483	--	--	487	491	493	--	--	497	742
9	9	9	9	6			8	8	6			8	5
(181 or 457)	458	475	479	483	--	--	487	491	493	--	--	497	742
(181 or 457)	458	475	479	483	--	--	487	491	493	--	--	497	742
9	9	9	9	6	482	486	8	8	6	494	499	497	742
9	9	9	9	6	8	7	8	8	6	8	9	497	742
(181 or 457)	458	475	479	483	--	--	487	491	493	--	--	497	742
9	9	9	9	6			8	8	6			8	5

Figure 2.

Pathways of allosteric communication between ATP and the lever arm for the *Dictyostelium* myosin II pre-stroke conformation. (a) Comparison of shortest paths derived from Identity, WAC, and Miyata matrices for the *Dictyostelium* pre-stroke structure 1VOM. The number underneath each residue (on a scale from 0 to 9) reflects the degree of relative conservation, with 9 corresponding to the highest conservation. Residues do not necessarily appear with the same conservation level in each row because of different weightings in the three matrices. Residues appearing in all three paths are darkest. (b) Visual representation of *Dictyostelium* myosin II pre-stroke pathway, computed using the Miyata matrix. The pathway residues (S181,

G457, F458, N475, Q479, F482, M486, F487, Q491, Y494, I499, E497, T742) are cyan. The path starts at the ATP γ -phosphate (green). The path ends at T742 of the lever arm (red).

Figure 3.

Alignment of paths computed from pre-stroke, post-rigor, and rigor structures. As explained in the text, all residue numbers are based on the alignment with *Dictyostelium* myosin II, which is used to define a common residue numbering system. The alignment is composed of 6 pre-stroke structures (light blue), 7 post-rigor structures (green), and 2 rigor structures (grey). Several residues are commonly shared across paths for different structures. In columns where significant residue conservation is observed within the pre-stroke or post-rigor conformation (measured as 50% or equivalently ≥ 3 of the sequences in this conformation), residues are colored according to their structural identity color. If such a conserved residue appears in another conformation, it is also colored in that conformation. When possible, corresponding residues are aligned in the same column. This is occasionally not possible due to a conflict in the ordering along the path. Residues in optimal paths (including ties) are connected by solid lines. Near-optimal paths (as defined in Methods) are indicated by dashed lines.

Figure 4.

Overlapped paths from pre-stroke and post-rigor structures. (a) and (b) *Dictyostelium* pre-stroke structure 1VOM, with the nucleotide (green) and terminal residue 742 (red) in space-filling representations, and path residues from 1VOM (teal), 1YV3 (blue), 1QVI (orange), 1LKX (purple), 1BR2 (pink) and 1W9J (brown) superimposed in stick representations. The relay helix is shown in cyan. Essentially the same paths were found in all six structures despite low sequence similarity amongst the four myosins represented. (c) and (d) *Dictyostelium* post-rigor structure 1MMD, with ADP-BeF₃ (green) and terminal residue 742 (red) in space-filling representations, and path residues from 1MMD (teal), 1MMA (blue), 1W7J (brown), 1W9I (orange), 1FMW (purple), 1KK7 (pink), and 1FMV (dark green). The relay helix is shown in cyan. Note the variation in paths close to the nucleotide in residues corresponding to the P-loop, switch I and switch II.

Figure 5.

Depiction of the *Dictyostelium* myosin II pre-stroke pathway residues 1) from structure 1VOM mapped onto the 1MMA post-rigor structure in light blue and 2) the actual pathway computed for 1MMA in yellow from the active site (green) to the beginning of the lever arm (red). The visible gaps in connectivity on the light blue path explain why the 1VOM pathway is not physically feasible in this post-rigor conformation and why the pathway computed for 1MMA (yellow) is so different.